

Carolin Müller-Spitzer

Der Einsatz einer maßgeschneiderten, feingranularen XML-Modellierung im lexikografischen Prozess

1. Zum Hintergrund dieses Beitrags

In *elexiko* sollen die Wortartikel so erarbeitet und kodiert werden, dass sie im elektronischen Medium flexibel dargestellt und abgefragt werden können. Die Entwicklung einer Modellierung, die diesen Ansprüchen genüge, nahm von Anfang an im Projekt eine wichtige Stellung ein, sollte sie doch die Voraussetzung dafür schaffen, dass ein innovatives Internetwörterbuch entwickelt werden konnte. Der vorliegende Beitrag soll die lexikografische Praxis im Vergleich mit der theoretischen Konzeption und den damit verbundenen Ansprüchen an die Modellierung beleuchten. Zum besseren Verständnis sollen dafür zunächst die grundlegenden Eigenschaften des Modellierungskonzeptes zusammengefasst werden.

Die in XML (eXtensible Markup Language, siehe unten ad 1) strukturierte Inhaltsstruktur in *elexiko* basiert auf einem Modellierungskonzept, das folgende grundlegende Eigenschaften aufweist:

- 1) Die Modellierung wird in XML-DTDs (Document Type Definitions, siehe unten ad 1) abgebildet.
- 2) Diese DTDs wurden maßgeschneidert für das Projekt *elexiko* entwickelt und entsprechen damit keiner Standard-Modellierung wie z. B. der TEI.¹
- 3) Die Modellierung ist streng inhaltsorientiert, d. h. sie klammert alle darstellungsbezogenen Informationen so weit als möglich aus.
- 4) Die Modellierung ist so feingranular wie möglich, d. h. kodiert alle lexikografischen (Vor-)Angaben einzeln.

¹ Die TEI (= Text Encoding Initiative) wurde Ende der 80er Jahre als Forschungsprojekt mit dem Ziel ins Leben gerufen, Richtlinien für die Auszeichnung verschiedener Texttypen aus dem geisteswissenschaftlichen Bereich zur Verfügung zu stellen. 1994 erschienen die „Guidelines for Electronic Text and Interchange“, bekannt als TEI P 3 (= TEI Proposal Number 3) (Sperberg-McQueen/Burnard 1995), worin auch ein Inventar für die Kodierung von Wörterbuchdaten enthalten ist. Da die TEI-Wörterbuchstruktur allerdings den Anspruch hat, für alle Wörterbücher der westlichen Welt anwendbar zu sein, ist sie für die meisten Projekte, die viele individuelle Zugriffsmöglichkeiten auf lexikografische Daten entwickeln wollen, zu allgemein gehalten.

- 5) Die DTDs sind in enger Zusammenarbeit zwischen den Verantwortlichen für die lexikografischen Inhalte einerseits sowie den Verantwortlichen für die XML-Struktur andererseits entwickelt worden.

Die Gründe für diese Ausrichtung der Modellierung sind folgende:

ad 1) Der Einsatz von XML bietet sich an, da XML ein softwareunabhängiger (Industrie-)Standard ist, der eine langlebige Datenhaltung garantiert und mit dessen Anwendung eine strikte Trennung von inhaltlicher Strukturierung und Präsentation erreicht werden kann.

The use of a single, internationally agreed syntax was not the only reason why a generic markup language was the appropriate solution for publishers that wanted to publish to multiple media. A fundamental concept in the use of generic markup languages is a move from formatting markup to content markup. (Gennusa 1999, S. 28)

Stefan Freisler spricht in diesem Zusammenhang auch von einer „Explizierung“ bzw. „Formalisierung der logischen Textstruktur“ (Freisler 1994, S. 41). Dabei kann XML als Metasprache gelten, mit der eigene Auszeichnungssprachen definiert werden können.

Zu XML gehört ein Grammatikformalismus, mit dem man in Form von so genannten *Document Type Definitions (DTDs)* kontextfreie Grammatiken [...] erstellen kann, um die Struktur einer Klasse von Dokumenten formal zu beschreiben. Eine DTD legt die erlaubten Tags fest und gibt Regeln an, wie sie verschachtelt werden dürfen. Dadurch kann mit einem Parser die *Gültigkeit* von Dokumenten bezüglich einer DTD-Grammatik festgestellt werden, um nicht nur die Wohlgeformtheit zu garantieren, sondern die Zuverlässigkeit zur Verarbeitung durch eine eingegrenzte Klasse von Softwaresystemen. (Hedtstück 2003, S. 132)

Mit der Bestimmung der Tags wird dabei in der DTD auch das Auszeichnungsvokabular festgelegt. In diesem Sinne ist eine DTD wiederum eine Metasprache zur Beschreibung einer Klasse von Dokumenten. Somit ist es mit XML möglich, eine individuell maßgeschneiderte Modellierung für ein konkretes lexikografisches Projekt zu entwickeln. Außerdem gibt es mittlerweile für XML zahlreiche flankierende Standards, wie die XML Path Language (XPath) oder XML Query Language (Xquery) zur Abfrage von XML-Dokumenten, oder die XML Stylesheet Language Transformations (XSLT) zur Transformation und damit auch zur Darstellung von XML-Dokumenten.

The key to understanding to revolutionary potential of XML is that it is just one piece of a larger picture. XML by itself can provide standardized interchange formats for databases and spreadsheets. This is significant. But XML and XSL together can replace existing word processing and desktop publishing formats as well. It can give us, in effect, a single, completely internationalized format

of almost unlimited power for both print and online publishing that is fully interoperable across all products and all platforms. The implications of this go far beyond data exchange and far beyond the Web. (Bosak 1998, S. 3)

ad 2) Eine maßgeschneiderte Modellierung erlaubt eine genaue Ausrichtung auf das jeweilige lexikografische Projekt. So kann sichergestellt werden, dass die lexikografische Mikrostrukturengrammatik so genau in eine XML-Modellierung überführt wird, dass sich ein Instruktionsbuch hinsichtlich der formalen Artikelstruktur in weiten Strecken erübrigt (vgl. Wiegand 1998, S. 232 und Müller-Spitzer 2007b, S. 70ff.). D. h., die Reihenfolge der Angaben etc. muss nicht von den Lexikografen selbst im Kopf behalten werden, sondern wird maschinell durch die Validierung im XML-Editor geprüft. Auch bestand die Annahme, dass die strenge, maßgeschneiderte Modellierung die Einarbeitung neuer Kollegen erleichtert. Darüber hinaus ist *elexiko* in ein Wörterbuchportal eingebunden: OWID,² das Online-Wortschatz-Informationssystem Deutsch des IDS. Die anderen Wörterbücher in OWID sind grundsätzlich nach dem gleichen Modellierungskonzept erarbeitet, sodass von der Modellierung her ein echtes Wörterbuchnetz geschaffen wurde, soweit dies die Inhalte zulassen (vgl. auch Müller-Spitzer 2007a). Sollte es im Kontext eines Projektverbundes sinnvoll und notwendig sein, könnten diese maßgeschneidert strukturierten Daten allerdings auch in ein standardkonformes Format wie z. B. die TEI migriert werden.

ad 3) Die Ausblendung jeder darstellungsbezogenen Information und stattdessen die Konzentration auf die Inhalte ist die Voraussetzung dafür, eine lexikografische Datenbasis zu erarbeiten, die für verschiedene Benutzungssituationen und -gruppen unterschiedlich darstellbar ist. Genau dies war für *elexiko* von Anfang an geplant.

ad 4) Auch die feingranulare Auszeichnung der lexikografischen Daten ist eine Voraussetzung für eine benutzeradaptive Darstellung. Außerdem soll so sichergestellt werden, dass alle Angaben potenziell recherchierbar sind und möglichst alle Teil-Ganzes-Beziehungen im Wortartikel maschinell kodiert sind, sodass interessante erweiterte Suchfunktionen entwickelt werden können (vgl. Müller-Spitzer 2007b, S. 124f.).

ad 5) Die XML-Modellierung als Teil der Kernaufgabe in einem lexikografischen Projekt zu begreifen, soll gewährleisten, dass die zu entwickelnde Struktur tatsächlich den Inhalten angemessen ist. Außerdem sind XML-DTDs als Notation auch für nicht in Modellierung geschulte Linguisten relativ leicht lesbar. Deshalb eignet sich XML prinzipiell gut als 'Kommunikationsformat'.

² Zu OWID vgl. <http://www.owid.de> und Müller-Spitzer (2010).

Die Syntax von XML-DTDs ist leicht zu verstehen, was ein elementarer Vorteil für die Strukturentwicklung ist. Gerade diese leichte Verständlichkeit ermöglicht die gemeinsame Diskussion über die zu entwickelnde Struktur unter allen Projektbeteiligten, auch und gerade mit denen, die für die Inhalte maßgeblich verantwortlich sind, für die die technische Umsetzung jedoch von geringem Interesse ist. Diese Kommunikation läuft nicht über Alltags- oder Fachsprache, sondern direkt anhand der Modellierung. Es ist in Projekten häufig zu beobachten, dass erst dann die Verständigung über die Modellierung detaillierter wird, wenn man gemeinsam den Entwurf einer DTD diskutiert. Erst die formale Syntax lässt es oft augenscheinlich werden, dass hier und da die entworfene Struktur doch den Inhalten nicht angemessen ist, dass es an bestimmten Stellen nicht berücksichtigte aber wohl begründete Ausnahmen gibt etc. (Müller-Spitzer 2007b, S. 86).

Um einen Eindruck zu bekommen, wie ein Wortartikel auf Ebene der Datenbasis in XML kodiert aussieht, ist im Folgenden ein Auszug des Wortartikels *abmahnen* zu sehen, der die Kodierung der Angaben zur Valenz illustriert (vgl. Abb. 1).

Über die theoretischen Grundlagen dieses Modellierungskonzeptes und seine Umsetzung wurde bereits Einiges publiziert (vgl. Müller-Spitzer 2005; 2007b, S. 243ff.; 2007c; 2007d). Hier soll es nun darum gehen, die praktischen Erfahrungen darzulegen, die alle Beteiligten im Team mit dieser Art der Modellierung gemacht haben. Um diese Erfahrungen detailliert nach verschiedenen Aspekten abzufragen, haben alle Kollegen, die Wortartikel in *ellexiko* verfassen, einen Fragebogen rund um ihre Erfahrung beim Arbeiten mit der Modellierung beantwortet. Viele Passagen in diesem Beitrag beruhen auf diesen Antworten.³ Sind im Text Zitate angeführt und nicht namentlich belegt, sind es Passagen aus den einzelnen Fragebögen. Im Fokus stehen dabei die Erfahrungen bei der Wortartikelerarbeitung, daneben werden aber auch Aspekte wie die Umarbeitung bereits geschriebener Artikel nach konzeptionellen inhaltlichen Änderungen, Erfahrungen bei der Weiterverarbeitung der Daten für die Darstellung sowie für die Entwicklung von Suchfunktionen ausgeführt. Insgesamt soll dieser Beitrag die lexikografische Praxis im Vergleich mit der oben zusammengefassten theoretischen Konzeption und den damit verbundenen Ansprüchen beleuchten und damit – wie die anderen Beiträge in diesem Band auch – einen Blick hinter die Kulissen bieten.

³ An dieser Stelle möchte ich mich dafür bei den *ellexiko*-Kollegen nochmals ausdrücklich bedanken.

```

<vb-valenz-neu>

<satzbauplan><satzbauplanA>
JEMAND mahnt JEMANDEN / ETWAS ab
</satzbauplanA></satzbauplan>
<satzbauplan><satzbauplanA>
JEMAND mahnt JEMANDEN / ETWAS (BEI / WEGEN ETWAS) ab
</satzbauplanA>
<angabe-zusatz>
<kommentar><lex-interpretationK>
<k-absatz>Zum obligatorischen Akkusativkomplement kann ein
Präpositivkomplement hinzukommen. Dabei sind die Präpositionen
<obj-text>bei</obj-text> und <obj-text>wegen</obj-text> im
lexiko-Korpus belegt.
<beleg>...</beleg></k-absatz></lex-interpretationK></kommentar>
</angabe-zusatz>
</satzbauplan>

<vb-komplemente-neu>
<subjekt-komp-neu obligatorisch="ja">
<nom-nominalphrase-neu/>
</subjekt-komp-neu>

<objekt-komp-vb obligatorisch="ja">
<akk-nominalphrase-vb/>
</objekt-komp-vb>

<objekt-komp-vb obligatorisch="nein">
<praepositionalphrase-vb praeposition="bei"/>
<praepositionalphrase-vb praeposition="andere"
andere="wegen"/>
</objekt-komp-vb>
</vb-komplemente-neu>

</vb-valenz-neu>

```

Abb. 1: Auszug der XML-Datei des *lexiko*-Artikels *abmahnen* mit den Angaben zur Valenz

2. Erfahrungen bei der Erarbeitung der XML-Struktur im lexikografischen Team

Der Prozess der DTD-Entwicklung hat sich in *lexiko* über einen langen Zeitraum erstreckt und lief damit Hand in Hand zur inhaltlichen Konzeptionsphase. Dadurch wurde das inhaltliche Konzept in den verschiedenen Angabebereichen immer parallel in eine formale Notation überführt, inhaltliche Änderungen, Nachbesserungen und auch erste Erfahrungen in der praktischen lexikografischen Arbeit mussten sich wiederum in den DTDs spiegeln. Im Nachhinein wird diese Vorgehensweise von allen noch im Projekt verbliebenen Kollegen, die seit Beginn der DTD-Entwicklung dabei waren, als guter und sinnvoller

Weg eingeschätzt. Bei der Befragung nach Vor- und Nachteilen dieser Vorgehensweise wurde vor allem die Rückkopplung zwischen inhaltlichem Konzept und Überführung in die DTD positiv hervorgehoben, gerade Überlegungen dazu, inwieweit bestimmte theoretische Konstrukte überhaupt in eine Struktur wie die einer DTD abbildbar sind. Als einziger Nachteil ist der Zeitaufwand zu nennen, den eine solche Vorgehensweise verlangt. Denn natürlich wurden in Tests immer wieder Mängel in den DTDs entdeckt, die jedoch oft nicht in der Modellierung selbst begründet waren, sondern in der Neuartigkeit der inhaltlichen Konzeption. Wie eine Kollegin es ausdrückt:

Mängel an der DTD haben sich dann später gezeigt, einerseits, weil die Konzeption nicht alles bedacht hatte und andererseits, weil die Praxis mit lexikografischen Daten zusätzliche Fälle oder anders geartete Fälle zum Vorschein brachte. Die Ursache für die meisten Nachbesserungen sehe ich hier: Die Konzeption beruhte hauptsächlich auf sprachtheoretischen, oftmals konventionellen, traditionellen sprachsystematischen Überlegungen, unsere Arbeit hingegen war dann am Sprachgebrauch orientiert. Das kontinuierliche Nachbessern war unumgänglich, war immer eingeplant und brachte stetig unsere eigene Erfahrung ein. Diesen ganzen Prozess sehe ich als einen an der Praxis ausgerichteten Prozess, den keiner zum Anfang hätte besser lösen können.

Voraussetzung für diese engmaschige Verbindung von Konzeption und Modellierung ist natürlich, dass alle Beteiligten eng kommunizieren und Verständnis für die unterschiedlichen inhaltlichen und informationstechnologischen Aufgaben entwickeln. Ist dies gegeben, entsteht ein interessanter Prozess, von dem beide – Lexikografen wie Informationstechnologen – profitieren können. Auf diese Weise kann daneben am besten sichergestellt werden, dass die Modellierung, auch wenn sie möglichst genau und streng die formale Artikelstruktur abbildet, keine unangenehme und unangemessene Gängelung der Lexikografen darstellt. Denn die Lexikografen sind selbst Teil der DTD-Entwicklungsgruppen und können damit von Anfang an großen Einfluss auf die Modellierung ausüben. „Unterm Strich erhält man eine ‘ideal’ auf die lexikografischen Erfordernisse ausgerichtete DTD, die man besser als eine vorge setzte verstehen und nachvollziehen kann“, wie eine Kollegin es ausdrückt. Insgesamt hat sich diese enge Kopplung von inhaltlicher lexikografischer Konzeption und Entwicklung der XML-basierten Inhaltsstrukturengrammatik (vgl. Müller-Spitzer 2007b, S. 129) für *elexiko* demnach als sinnvoll erwiesen.

Die andere Frage ist die, ob sich die XML-DTDs im Nachhinein wirklich als ‘Schnittstellenformat’ bewährt haben oder nicht. Wie oben gesagt, war ein Grund für die gemeinsame DTD-Entwicklung die These, dass XML-DTDs ein geeignetes Kommunikationsformat zwischen Geisteswissenschaft und Informatik sind. Dazu ist allerdings zu sagen, dass in der Praxis wenige Kollegen

die Modellierung wirklich nur durch das Lesen der entworfenen DTD geprüft haben, dafür war die Syntax, obwohl an sich nicht komplex, doch zu „gewöhnungsbedürftig“. Vermehrt wurde deshalb die entworfene Struktur dadurch getestet, dass die Test-DTD in einen XML-Editor geladen wurde und so ein Teilbaum der künftigen Artikelstruktur ausgetestet wurde. Allerdings resümiert eine Kollegin auch:

Die XML-Technologie ist für mich gut verständlich, und ich hätte keinesfalls weniger Einblick in die Technik haben wollen. Ich finde es zum besseren Verständnis der Modellierung für den Lexikografen unabdingbar, dass er die DTD mindestens einmal zu Gesicht bekommen hat. Für mich war und ist es von Vorteil, viel von der DTD-Entwicklung mitzubekommen, weil mir das jetzt, wenn ich mir neue Angaben oder Änderungen überlege, dabei hilft, gleich zu überlegen, wo man das wie in der DTD einbauen könnte. Ich kann jetzt besser abschätzen, was modellertechnisch möglich ist und was nicht.

Insgesamt kann man zusammenfassen, dass die Diskussion eines Modellierungsentwurfes nicht – wie vielleicht erwartet oder erhofft – hauptsächlich anhand der formalen Notation der DTD geführt wurde, weil die Notation dafür manchen zu ungewohnt ist. Im Nachhinein empfanden es aber alle als hilfreich, dass sie sich durch das Beschäftigen mit XML und durch die regelmäßigen Tests der entworfenen DTDs mehr mit der Technologie befasst haben und sich so diese Technik zu eigen gemacht haben. Letzteres ist für ein Projekt wie *ellexiko* eigentlich unerlässlich, denn je mehr die Lexikografen die XML-Modellierung als Fremdkörper empfinden, desto mehr besteht die Gefahr, dass sie sich in der Artikellarbeit von den ursprünglichen Ideen, die zu einer bestimmten Modellierung geführt haben, entfernen. Haben sie sich mit den Intentionen der Modellierung vertraut gemacht, werden die Inhalte mancher Auszeichnungen zwar trotzdem im Laufe der Artikellarbeit unbemerkt nach und nach verändert, aber doch in geringerem Maße, als dies möglicherweise sonst der Fall wäre. Bei *ellexiko* hat sich vor allem gezeigt, dass diese Entfremdung von ursprünglich etwas anders gedachten Angabetypen sich vor allem in einem Bereich verselbstständigt hat, für den keiner speziell die inhaltliche Verantwortung trug, nämlich im Bereich der Angabezusätze wie Belege, Hinweise oder Kommentare. In allen anderen Bereichen, in denen jeweils einer aus der Projektgruppe gemeinsam mit der informationstechnologischen Seite die Verantwortung getragen hat, ist dies sehr viel weniger zu beobachten, da insbesondere die inhaltliche Verantwortlichkeit für einen Angabebereich auch bedeutet, die lexikografische Praxis hinsichtlich dieser Angaben besonders im Auge zu behalten. Auch das zeigt, dass sich das in *ellexiko* gewählte Modell bei der Entwicklung der XML-Struktur zumindest für eine solch reiche Mikrostruktur, wie sie ein bearbeiteter *ellexiko*-Artikel hat, bewährt hat.

3. Erfahrungen bei der Einarbeitung neuer Kollegen

Ein Grund, weshalb in *elexiko* eine maßgeschneiderte, strenge Modellierung gewählt wurde, war die Erwartung, dass eine solche Modellierung die Einarbeitung neuer Kollegen erleichtern würde und somit für ein langfristiges Projekt mit wechselnder personeller Besetzung von Vorteil ist. Deshalb wurden bei der Befragung nach den Erfahrungen mit der XML-Modellierung diejenigen Kollegen, die später zum Projektteam hinzugekommen sind, explizit nach ihren Erfahrungen bei der Einarbeitung hinsichtlich der XML-Struktur befragt. Hilfreich fanden diese Kollegen vor allem, dass sie durch die Modellierung in der Einhaltung der formalen Artikelstruktur stark unterstützt wurden und so „die inhaltlichen Aspekte bei der Einarbeitung im Vordergrund stehen“ konnten. Auch wurde die explizite Benennung der XML-Tags (vgl. Müller-Spitzer 2005, S. 32ff.) als Unterstützung empfunden:

Hilfreich fand ich vor allem, dass die Tags [Auszeichnungen, Anm.d.Verf.] oft sprechende Namen haben und sich nach einem bestimmten System richten – beispielsweise, dass der Begriff *Angabe* mit einem großen A abgekürzt wird.

In dieser Hinsicht haben sich die Anfangserwartungen damit erfüllt.

Schwierigkeiten gab es rückblickend kaum im Bereich der eigentlichen Modellierung, sondern mehr in der Arbeit mit einem standardisierten, nicht an die Projektbedürfnisse angepassten XML-Editor. Zwar fanden auch die neuen Kollegen die Arbeit in der Tag-Ansicht nicht schwierig, sehr wohl aber z. B. den Umgang mit Fehlern in der Struktur, die bei der Validierung aufgezählt werden, oder die Unterscheidung, welche Elemente der Wortartikelstruktur obligatorisch und welche fakultativ sind. Hier bestanden Schwierigkeiten darin, die angemarkten Fehler auch wirklich zu finden (z. B. unterscheiden zu lernen, welche Fehlermeldungen sich auf Elemente oder Informationen über Elemente, in XML Attribute genannt, beziehen) und sie entsprechend zu beheben.

Auf die Frage, was man rund um die XML-Struktur für die Einarbeitung verbessern könne, wurden dementsprechend eher Wünsche geäußert, die nicht direkt etwas mit der XML-Struktur für *elexiko* zu tun haben, sondern mehr mit der Anpassbarkeit des XML-Editors (z. B. visuell deutlicher zwischen obligatorischen und fakultativen Elementen zu unterscheiden) oder mit Informationen rund um die XML-Struktur, z. B. die inhaltlichen Richtlinien stärker im Redaktionshandbuch zu dokumentieren.

Eine weitere Frage in Hinsicht auf die Einarbeitung war auf die Benutzung dieses Redaktionshandbuches gerichtet. Das Redaktionshandbuch von *elexiko* ist von der Organisation her eine DTD-Dokumentation, die mit Hilfe der Soft-

ware DTDhelp automatisch aus der DTD generiert wird, d.h. aus der XML-DTD wird durch das Tool eine DTD-Dokumentation in Form einer XML-Datei generiert. Diese Dokumentation kann in zwei Formaten (zur internen Benutzung) ‘publiziert’ werden: als HTML- oder RTF-Format. In *ellexiko* wird ausschließlich die HTML-Version der Modellierung genutzt, zum einen, weil sie als kontextsensitive Hilfe in den XML-Editor eingebunden und so beim Wortartikelschreiben komfortabel aufgerufen werden kann, zum anderen, weil das

OVIDIUS

➤

EWL-ARTIKEL.DTD
Alle Elemente
<form>

Modellierung

Form

Hier kann die angegebene Form durch eine Verwendungshäufigkeitsangabe ergänzt werden. Die Klammerung mit dem Element `form` ist notwendig, da sonst nicht festgelegt werden kann, dass erst die Form-Angabe zu erfolgen hat und dann die Kommentierung o.ä. Verwendet in der **Grammatik** und in lesartenübergreifenden Angaben.

Redaktionshandbuch

Immer wenn es sich um schwankende **Formen** handelt, müssen diese Formen durch eine Verwendungshäufigkeitsangabe spezifiziert werden.

Getaggte Verwendungshäufigkeitsangaben (z.B. bei Varianten des Genitiv Singulars) müssen nur dann angegeben werden, wenn tatsächlich beide Formen im *ellexiko*-Korpus belegt sind.

Beispiel:

```

<angabe-zusatz>
  <kommentar>
    <lex-interpretationK>
      <k-absatz>Präpositionalkomplemente mit der Präposition
      <hervorhebung typ="kursiv">bei</hervorhebung> sind selten im
      Korpus belegt. Vgl.
    <belege>
      <ek-beleg><zeitung-beleg><belegtextA>Noch immer wird dennoch
      bei Straßenplanungen mit Gelände
      <belegwortA>geaast</belegwortA></belegtextA>
      <zt-belegnachweisA name="MM" ressort="Politik"
      datierung="18.04.1987">Mannheimer Morgen, 18.04.87, S. 2,
      Osterspaziergang.
      </zt-belegnachweisA></zeitung-beleg>
    </ek-beleg>
  </belege>
</k-absatz>
</lex-interpretationK>
</kommentar>
</angabe-zusatz>

```

Gibt es mehrere **Formvarianten**, erst die häufigeren übernehmen.

Kontext-Diagramm

Ist enthalten in

akk-pl | akk-sg | dat-pl | dat-sg | gen-pl | gen-sg | imp-sg23 | komparativ | konj-praet | morpho-variante | nm-nom-pl | nom-pl | nom-sg | ortho | ortho-variante | partI | partII | praet | sg123 | sg23 | superlativ

Enthält

angabe-zusatz | formA | verw-hkeitA

Abb. 2: Eintrag zum Element `form` aus dem *ellexiko*-Redaktionshandbuch

Redaktionshandbuch kontinuierlich aktualisiert wird und ein immer neuer Ausdruck von aktuell etwa 500 Seiten Dokumentation alles andere als sinnvoll erscheint. Als feste Bestandteile finden sich in diesem Redaktionshandbuch zu jedem Element aus der DTD ein Namensfeld, ein Navigationsfeld, über das zu Ober- bzw. Unterelementen navigiert werden kann, sowie weitere frei zu definierende Felder. In *ellexiko* sind dies immer ein Feld zur Kommentierung der Modellierung (falls nötig) sowie ein Feld, in dem inhaltliche Richtlinien festgehalten werden (vgl. Abb. 2). Die inhaltliche Füllung des Redaktionshandbuchs nimmt dabei die meiste Zeit in Anspruch.

Das Redaktionshandbuch wird jedoch von den meisten Kollegen im Projekt nicht als DTD-Dokumentation verwendet, sondern vor allem in Hinsicht auf die inhaltlich-redaktionellen Richtlinien, die bestimmten DTD-Elementen zugeordnet sind. Die eigentliche DTD-Dokumentation ist nur für diejenigen nützlich, die selbst an der Modellierung arbeiten, sowie für die Formulierung von umfangreichen erweiterten Suchen mittels XPath, um den entsprechenden Pfad herauszusuchen (vgl. Abschnitt 7).

Insgesamt kann man also sagen, dass sich die strenge, maßgeschneiderte und kontinuierlich angepasste Modellierung für die Einarbeitung neuer Kollegen gut bewährt hat. Allerdings muss betont werden, dass sich diese Erleichterung nur auf die formalen Aspekte der Artikelstruktur beschränkt. Inhaltliche Richtlinien zur Füllung dieser Struktur kann keine noch so strenge Modellierung transportieren. Die Erarbeitung dieses Wissens kann durch die Art der Modellierung daher kaum unterstützt werden.

4. Erfahrungen bei der Wortartikelerarbeitung

4.1 Vor- und Nachteile bei der täglichen Artikelarbeit

Entgegen der oft formulierten Annahme ‘klassischer’ Lexikografen, eine strenge Struktur würde die Arbeit an einem Wortartikel zu stark gängeln und die Mitarbeiter einengen, empfinden die *ellexiko*-Lexikografen die sehr umfangreiche, granulare und strenge XML-Struktur laut ihren Antworten im Fragebogen gerade als Vorteil. So schreibt eine Kollegin auf die Frage, worin sie einen Vorteil beim Arbeiten in der XML-Struktur sieht:

Struktur und Gliederung – man hat ‘Stationen’, an denen man geschickterweise unterbrechen und wieder beginnen kann. Da lexikografische Arbeit immer auch einen Anteil (lästiger) Routine mit sich bringt, vergisst man deutlich weniger: Man wird geführt, auch mal erinnert. Die Informationsdichte/-verteilung bleibt in einer ausgewogenen Relation – innerhalb der eigenen Artikel

und auch im Verhältnis der einzelnen Autoren zueinander. Die klare Unterscheidung in einzelne Informationsbereiche hilft auch beim konzeptionellen Gliedern der Daten/Informationen eines Gesamtartikels – und letztlich auch beim methodischen Aufarbeiten der Korpusanfragen.

Auch die Möglichkeit der relativ detaillierten Validierung, die eine so feingranulare Modellierung ermöglicht, wird als Vorteil gesehen:

Die Option der Validierung finde ich besonders wichtig und schön, da man wirklich nichts vergessen kann, denn das würde mir regelmäßig passieren, mit ein paar kleineren Attributen vor allem. Wenn ich am Ende validiere, gibt es immer etwas, das fehlt.

Die DTD bzw. ihre Umsetzung im XML-Editor führt einen an vielen Stellen durch die vorgesehene Struktur. Obligatorische Angaben kann man nicht vergessen. Durch die Auswahlmöglichkeit bei Attributwerten ist Verschreibungen (nicht aber Verklickungen!) vorgebeugt. Die expliziten Tag-Benennungen helfen dabei zu verstehen, was man wo eintragen soll.

Gefragt nach Nachteilen, konnten die Lexikografen erfreulicherweise wenig anführen. Als einziger Nachteil wird gesehen, dass man bestimmte Korpusbefunde nicht spontan eintragen kann, sondern erst, wenn man sich an die entsprechende Stelle in der Artikelstruktur geklickt hat. Diesem Bedürfnis müsste allerdings mehr durch eine Ergänzung der Redaktionsumgebung Rechnung getragen werden, indem z. B. eine Art Notizfunktion für einen Wortartikel eingeführt wird, in die zunächst unstrukturiert Ideen oder Korpusbefunde etc. eingetragen werden können. Als lästig werden weiterhin bestimmte (fehlende) Komponenten in der Redaktionsumgebung empfunden, wie z. B., dass es nicht möglich ist, einen im XML-Editor verfassten Artikel direkt als Online-Preview anzusehen. An solchen Verbesserungen der redaktionellen Umgebung wird allerdings im Projekt *BZVlexiko*⁴ gerade gearbeitet.

In einem standardisierten XML-Editor zu arbeiten stört (auch entgegen häufig formulierter Bedenken externer Kollegen) die beteiligten Lexikografen prinzipiell wenig. Auch in diesem Bereich ist es die Aufgabe des *BZVlexiko*-Projektes, die pure und standardisierte Funktionalität des eingesetzten XML-Editors komfortabler für die Lexikografen zu gestalten (z. B. zur Verwaltung der Vernetzungen, dem Einsatz von Templates für bestimmte Angabebereiche etc.), u. a. um die oben geschilderten, eher unkomfortablen Eigenschaften der Redaktionsumgebung zu verbessern.

⁴ Zum Projekt *BZVlexiko* vgl. das Vorwort in diesem Band.

4.2 Granularität der Struktur

Wie in Abschnitt 4.1 bereits beschrieben wurde, empfinden die Lexikografen die Granularität der *elexiko*-Modellierung als vorteilhaft bei der Artikellarbeit in der Hinsicht, dass keine Angaben vergessen werden können und die formale Konsistenz der Wortartikel durch die Validierung sichergestellt wird. Allerdings spiegelt die *elexiko*-Modellierung die theoretische Konzeption und damit auch Annahmen über die Struktur sprachlicher Befunde wider, die sich im Laufe der korpusbasierten Arbeit nicht immer so bestätigt haben. Dieses Thema – theoretische Vorannahmen im Verhältnis zu den empirisch gewonnenen Befunden – zieht sich durch alle Beiträge dieses Sammelbandes und spielt auch im Kontext der Modellierung eine große Rolle. Ein kleines Beispiel dazu: Bei der Grammatik von Nomina ist vorgesehen, dass zur Singular- und Pluralverwendung die Angaben ‘Singular+Plural’, ‘nur Singular’ oder ‘nur Plural’ gegeben werden können. Dazu sagt die für die Grammatik verantwortliche Kollegin heute:

Wir haben fast keine Fälle, die in die harten Abweichungskategorien passen (bei Pluraletantum z. B. nur *Eltern* oder *Ferien*) und aber einige Fälle, bei denen man korpusgestützt gerne eine gewisse Auffälligkeit beschreiben würde. Deshalb wäre es vielleicht besser gewesen zu sagen: Singular+Plural ist das Normale. Das Normale muss aber nicht obligatorisch ausgefüllt werden, sondern es gibt auch die Möglichkeit, einen freieren Text einzugeben (z. B. „meist im Plural“, „auffällig oft im Singular“, „immer nur Plural“).

Vor diesem Hintergrund wäre es angebracht gewesen, bestimmte Kategorien, die sich im Laufe der empirischen Arbeit als nicht sinnvoll erwiesen haben, im Nachhinein zu modifizieren und durch freiere, dem tatsächlichen Sprachgebrauch angemessene Kategorisierungen zu ersetzen. In einem Wörterbuchprojekt wie *elexiko* bedeutet dies jedoch immer einen nicht zu unterschätzenden Aufwand für die nachträgliche Umarbeitung bereits geschriebener Artikel.

5. Erfahrungen bei der nachträglichen Umarbeitung der Wortartikel

elexiko ist ein Wörterbuchprojekt, das kontinuierlich weiterentwickelt wird, nicht nur in Bezug auf den Umfang, sondern auch konzeptionell. Dies hat zur Folge, dass die dem *elexiko*-Demonstrationswortschatz⁵ zugrunde gelegte Artikelstruktur hinsichtlich mancher Angabebereiche erhebliche Änderungen erfahren hat, z. B. im Bereich der typischen Verwendungsmuster (vgl. Storjohann/Möhrs 2007 und den Beitrag „Die typischen Verwendungen in *elexiko*“ von Christine Möhrs in diesem Band), in der Valenz oder in der Paradigma-

⁵ Vgl. hierzu Abschnitt 3 der Einleitung in diesem Band.

tik (vgl. insgesamt Hahn et al. 2008). Dabei war es immer der Anspruch von *ellexiko*, alle bearbeiteten Artikel auf einem gleichen inhaltlichen Stand zu halten, d. h. nicht nur die nach einer konzeptionellen Änderung neu erarbeiteten Artikel entsprechend der neuen Vorgaben zu schreiben, sondern auch die bereits verfassten entsprechend nachzubearbeiten. Dies hat vor allem inhaltliche Gründe, nämlich dass Benutzer eine konsistente Datenstruktur bei allen bearbeiteten Wortartikeln vorfinden sollen. Daneben gibt es auch eher technische Gründe, die für diese Umarbeitungen sprechen: Würden die bereits verfassten Wortartikel nicht in die neue Struktur gebracht, würden sie auch nicht der gleichen DTD gehorchen und dementsprechend auch nicht mit dem gleichen Stylesheet anzuzeigen sein. Diese Heterogenität zu verwalten wäre nicht ganz unproblematisch.

Solche Nachbearbeitungsschritte sind in *ellexiko* allerdings mit einem hohen Arbeitsaufwand verbunden, weil viele Schritte vollzogen werden müssen:

- Die DTD muss der neuen Konzeption angepasst werden.
- Die neue DTD muss getestet werden.
- Bereits fertige Wortartikel müssen inhaltlich der veränderten Konzeption angepasst werden.
- Das Stylesheet muss der veränderten DTD angeglichen werden (vgl. ausführlicher Hahn et al., S. 78ff.).

Insbesondere der letzte Schritt war bei einigen Änderungen, z. B. bei den typischen Verwendungen und bei veränderten inhaltlichen Festlegungen in der Paradigmatik, sehr aufwendig. Zwar können manche Veränderungen rein in der XML-Struktur durch automatische Konvertierung aufgefangen werden, aber das eigentlich Aufwendige, nämlich die inhaltliche Nachbearbeitung der Wortartikel, kann kaum automatisch unterstützt werden. Darüber hinaus ist an solchen Punkten die schlichte Redaktionsumgebung von *ellexiko* von großem Nachteil: Es können bislang weder Artikelstrecken noch partielle Wortartikel aus der Datenbank ausgecheckt (d. h. zur Bearbeitung entnommen) werden. Jeder Artikel muss einzeln ausgecheckt, im XML-Editor geöffnet, validiert und wieder eingeecheckt werden. Hier zeigt sich auch ein Nachteil der granularen XML-Struktur: Fast jede inhaltliche Änderung hat auch eine DTD-Änderung und dementsprechend eine Stylesheetänderung zur Folge.

Im Nachhinein muss man daher konstatieren: Es wäre sinnvoll gewesen, nach der Erarbeitung des *ellexiko*-Demonstrationswortschatzes eine noch deutlichere Zäsur zu machen und zu diesem Zeitpunkt vor dem Hintergrund der bis dahin erfolgten lexikografischen Arbeit die XML-Struktur einer grundlegenden Revi-

sion zu unterziehen. Manche konzeptionellen Änderungen hätten zu diesem Zeitpunkt noch wesentlich weniger Aufwand an Nachbearbeitung erfordert als zu einem späteren Zeitpunkt. Auf der anderen Seite bleibt es bei einem Projekt wie *elexiko*, das den Anspruch hat, lexikologische und metalexikografische Forschungsergebnisse möglichst schnell in die lexikografische Arbeit einfließen zu lassen, nicht aus, auch im laufenden Wörterbuchbetrieb Änderungen an der Wortartikelstruktur vorzunehmen und die entsprechend notwendigen Änderungen an bereits geschriebenen Artikeln in Kauf zu nehmen. Der Arbeitsaufwand für diese Umarbeitungen ist jedoch erheblich.

6. Erfahrungen bei der Entwicklung der XSLT-Stylesheets

Sehr viel aufwendiger als am Anfang gedacht war auch die Entwicklung des XSLT-Stylesheets. Bei der Entwicklung der Modellierung war zunächst nur die lexikologisch-lexikografische Konzeption von Interesse – Fragen der Darstellung blieben weitestgehend ausgeklammert. So wurde auch die Reihenfolge der Angaben in den DTDs nur danach ausgerichtet, was inhaltlich zusammengehört und von der Bearbeitung her sinnvoll zueinander passt. Wie im ersten Kapitel in den Grundlagen der Modellierung beschrieben, war dabei die Vorgabe, alle Angaben separat auszuzeichnen, um sie recherchierbar und adaptiv darstellbar zu machen. Die *elexiko*-Wortartikelstruktur wurde als DTD damit sehr umfangreich: Sie umfasst etwa 400 Elemente und 300 Attribute, für die alle eine Behandlung im Stylesheet definiert werden muss. Dabei wurde zunächst (und bis heute) eine Standardansicht entwickelt. Hätte man von vornherein eine benutzeradaptive Ansicht entwickeln wollen, d. h. unterschiedliche Ansichten der gleichen XML-Elemente und Attribute je nach Benutzungssituation und Benutzergruppe, hätte sich dieser Aufwand potenziert. Abgesehen von dem Aufwand fehlte es bislang an empirischen Erkenntnissen über die Benutzung von Wörterbüchern im Bereich der Online-Lexikografie, die jetzt allerdings im BZ*Velexiko*-Projekt in umfangreichen empirischen Studien erhoben werden.

Da zum *elexiko*-Team bedauerlicherweise kein Informatiker zählte, wurden die Stylesheets extern entwickelt. Dies war zu diesem Zeitpunkt die pragmatischste Lösung, barg langfristig jedoch einige Schwierigkeiten: Beispielsweise konnte die DTD-Entwicklung nicht hinreichend mit der Stylesheetentwicklung abgestimmt werden. Deshalb finden sich in den DTDs Konstruktionen, die mit der XSLT-Technologie schwierig umzusetzen sind.

Ein Beispiel für ein solches Element ist der so genannte Angabe-Zusatz. Dieses Element enthält Belege, Hinweise und Kommentare. Diese werden nicht an

einer Stelle in der Artikelstruktur gesammelt, sondern immer explizit einzelnen Angaben oder Angabegruppen zugeordnet. Auch dies war die Einlösung eines in der metalexikografischen Forschung formulierten Desiderats, nämlich dass Belege und andere Zusätze zu Angaben auch von ihrem Skopus her genau diesen Angaben zugeordnet sein sollten (vgl. Müller-Spitzer 2007b, S. 152ff.).

Dieser Ansatz führte allerdings dazu, dass in der *elexiko*-DTD an extrem vielen Stellen ein Angabe-Zusatz vorgesehen ist, da man theoretisch an allen möglichen Stellen das Bedürfnis haben könnte, eine Angabe oder Angabegruppe zu kommentieren, einen Hinweis zu geben etc. Da alle diese Zusätze in der DTD zu einem Element zusammengefasst sind (Belege, Hinweise und Kommentare) führt dies zu einer eigentlich nicht erwünschten Aufweichung der strengen Struktur. Denn in diesen Zusätzen können beliebig viele Kombinationen von Belegen und Hinweisen bzw. Kommentaren gegeben werden, d. h. teilweise textuell lange Angaben. Hier wäre es eine bessere Strategie gewesen, nur an den Stellen, an denen in der praktischen Artikellarbeit wirklich das Bedürfnis auftritt, eine Angabe zu kommentieren, zu belegen oder einen Hinweis zu geben, diese Möglichkeit vorzusehen, anstatt sie an allen theoretisch denkbaren Stellen präventiv einzubauen. Abgesehen davon stellen diese Angabe-Zusätze an hierarchisch ganz unterschiedlichen Orten ein erhebliches Problem bei der Stylesheeterstellung dar, da sie an unterschiedlichen Stellen in sehr unterschiedlicher Länge mit z. T. unterschiedlichen Funktionen auftauchen. Wäre dies aufgrund eines Austauschs im Team bekannt gewesen, wäre die Modellierung der Angabe-Zusätze rechtzeitig geändert und so entsprechende Schwierigkeiten vermieden worden.

Auch in anderer Hinsicht wäre es besser gewesen, die Stylesheetentwickler in das Projektteam zu integrieren. Es wäre einfacher gewesen, Änderungswünsche direkt zu besprechen, es wäre mehr Kontinuität gewährleistet gewesen und solche Entwickler hätten sich auch um andere eher technologische Belange kümmern können, die jetzt – zu einem wesentlich späteren Zeitpunkt – im BZ*Velexiko*-Projekt beispielsweise mit der Entwicklung eines Vernetzungsmanagers realisiert werden. Dies ist demnach auch eine Lehre: Man kann ein Online-Wörterbuch mit einer so ambitionierten und aufwendigen Form der Datenmodellierung, ohne entsprechend fähige technische Mitarbeiter, die für alle Belange rund um diese Datenmodellierung – nötige Konvertierungen, Entwicklung der Stylesheets, Verbesserung der Redaktionsumgebung – zuständig sind, nicht gut in die Praxis umsetzen.

7. Erfahrungen bei der Entwicklung von Suchfunktionen

Wie anfangs erläutert, war der Grund für die feingranulare Modellierung auch der Wunsch, möglichst viele erweiterte Suchmöglichkeiten entwickeln zu können. Intern ist dies mit der Umstellung auf das neue Datenbanksystem EDAS (Electronic Dictionary Administration System, vgl. Müller-Spitzer/Schneider 2009) inzwischen möglich. Die Lexikografen können selbst mittels XPath detaillierte Anfragen über die bearbeiteten Artikel stellen. Gesucht werden kann nach allen Elementen und allen Attributen, auch in kombinierter Form und mit bestimmten Inhaltsbeschränkungen. Ein Beispiel für eine relativ schlichte erweiterte Suche ist z. B. die nach allen Artikeln, in denen in der Bedeutungsparaphrase „Computer*“ vorkommt. Diese Anfrage zeigt als Suchergebnis Stichwörter wie „mobil“, „Endanwender“, „Netz“ oder „Platte“ an. Spezialisiertere Anfragen sind z. B. solche nach allen Verben, die ein nicht-obligatorisches Objekt-Komplement haben, welches als Präpositionalphrase mit der Präposition *an* realisiert wird (vgl. Abb. 3).

Erweiterte Suche

enthält	//ewl-artikel//vb-komplemente-neu	mit Inhalt
enthält	//objekt-komp-vb[@obligatorisch="nein"]	mit Inhalt
[entfernen]		
enthält	//praepositionalphrase-vb[@praeposition="an"]	mit Inhalt
[entfernen]		

Suche starten

1. fehlen
2. mangeln
3. scheitern
4. weben
5. zunehmen
6. zweifeln

Suche beendet um 15:29:11h.

Abb. 3: Beispiel einer erweiterten Suche in EDAS

Auf der Benutzeroberfläche kommt dieses Potenzial der feingranularen Modellierung allerdings noch nicht zur Geltung. Zwar gibt es erweiterte Suchen zu *ellexiko*, diese sind aber sehr beschränkt und bieten den Benutzern nicht annähernd die Zugriffsstrukturen, die von der XML-Modellierung möglich und auch für die Wörterbuchbenutzung sinnvoll wären. Auch hier hat sich gezeigt, dass die mangelnde technische Kompetenz im *ellexiko*-Team von großem Nachteil war. Denn es ist eine Sache, die Voraussetzungen für detaillierte Suchanfragen auf Seiten der Modellierung und Datenauszeichnung zu

schaffen, wie es in *ellexiko* gemacht wurde, aber eine andere Sache, diese Voraussetzungen und Möglichkeiten auch in Funktionalitäten für Benutzer umzusetzen. Letzteres setzt voraus, eine Art der Speicherung und eine Art des Retrievals der XML-Instanzen in der Datenbank zu entwickeln, die detaillierten und trotzdem noch performanten Zugriff auf die XML-Instanzen erlaubt. Allerdings werden auch hier aufgrund besserer personeller Ausstattung in absehbarer Zeit Verbesserungen vorgenommen werden können.

8. Resümee

Der Einsatz einer feingranularen, maßgeschneiderten Modellierung hat sich insgesamt für *ellexiko* bewährt. Die Lexikografen kommen bei der Wortartikel-erarbeitung gut mit der Struktur zurecht, sie erleichtert ihnen das Arbeiten, soweit das eine formale Struktur kann, sie hilft bei der Einarbeitung neuer Kollegen und schafft die Voraussetzungen für einen flexiblen Umgang mit lexikografischen Daten, z. B. für einen benutzeradaptiven Zugang zum Wörterbuch. Sollte es im Austausch mit anderen Projekten nötig sein, könnte diese maßgeschneiderte Modellierung jederzeit in ein standardkonformes Format wie die TEI überführt werden.

Am Anfang zu wenig beachtet wurde in *ellexiko* der technologische Aufwand, der mit dem Einsatz einer solchen Modellierung verbunden ist. Hätte das Projektteam von Anfang an einen Informatiker als Mitglied gehabt, hätten sicherlich manche Arbeitsprozesse, die viel Zeit gekostet haben, optimiert werden können. Doch damit teilt *ellexiko* sicher das Schicksal vieler Projekte, die einen neuen Weg zu gehen wagen. Außerdem musste an einem Institut wie dem IDS wie sicher auch in vielen anderen geisteswissenschaftlich ausgerichteten Instituten ein Prozess des Umdenkens stattfinden in der Hinsicht, dass die Erstellung eines Online-Wörterbuchs auch ein technologisches Projekt ist, in dem die technologischen Arbeiten nicht nur ein peripherer und punktueller Bestandteil der Arbeit, sondern ein substanzieller Teil davon sind. Hat dieser Prozess stattgefunden, können auch die entsprechenden Konsequenzen gezogen werden, nämlich ein Projektteam zusammenzustellen, das nicht ausschließlich aus Linguisten besteht, sondern genauso Informatiker oder Computerlinguisten umfasst. Das Zusammenspiel dieser unterschiedlichen Kompetenzen und Herangehensweisen ist dann wiederum ein Lernprozess, der für beide Seiten spannend und bereichernd ist und dem gemeinsam erarbeiteten Produkt unbedingt zugute kommt.

9. Literaturverzeichnis

9.1 Forschungsliteratur

- Bosak, Jon (1998): Media-independent publishing: four myths about XML. Internet: www.ibiblio.org/pub/sun-info/standards/xml/why/4myths.htm (Stand: 30.04.2010).
- Freisler, Stefan (1994): Hypertext – Eine Begriffsbestimmung. In: Deutsche Sprache 22, S. 19-50.
- Gennusa, Pamela L. (1999): Evolution and use of generic markup languages. In: Möhr, Wiebke/Schmidt, Ingrid (Hg.): SGML und XML. Anwendungen und Perspektiven. Berlin/Heidelberg/New York, S. 27-50.
- Hahn, Marion/Klosa, Annette/Müller-Spitzer, Carolin/Schnörch, Ulrich/Storjohann, Petra (2008): *elexiko* – das elektronische, lexikografisch-lexikologische korpusbasierte Wortschatzinformationssystem. Zur Neukonzeption, Erweiterung und Revision einzelner Angabebereiche. In: Klosa, Annette (Hg.): Lexikografische Portale im Internet. (= OPAL – Online publizierte Arbeiten zur Linguistik 1/2008). Mannheim, S. 57-85. Internet: www.ids-mannheim.de/pub/laufend/opal/privat/pdf/opal08-1_hahn-klosa-mueller-spitzer.pdf (Stand: 18.05.2010).
- Hedstüch, Ulrich (2003): Einführung in die theoretische Informatik. Formale Sprachen und Automatentheorie. 2., überarb. Aufl. München/Wien.
- Müller-Spitzer, Carolin (2005): Die Modellierung lexikografischer Daten und ihre Rolle im lexikografischen Prozess. In: Haß, Ulrike (Hg.): Grundfragen der elektronischen Lexikographie. *elexiko* – das Online-Informationssystem zum deutschen Wortschatz. (= Schriften des Instituts für Deutsche Sprache 12). Berlin/New York, S. 20-54.
- Müller-Spitzer, Carolin (2007a): Das *elexiko*-Portal: Ein neuer Zugang zu lexikografischen Arbeiten am Institut für Deutsche Sprache. In: Rehm, Georg et al. (Hg.): Datenstrukturen für linguistische Ressourcen und ihre Anwendungen. Proceedings of the Biennial GLDV Conference 2007, 11.-13. April, 2007. Tübingen, S. 179-188.
- Müller-Spitzer, Carolin (2007b): Der lexikografische Prozess. Konzeption für die Modellierung der Datenbasis. (= Studien zur Deutschen Sprache 42). Tübingen.
- Müller-Spitzer, Carolin (2007c): Das Konzept der Inhaltsstruktur. Eine Auseinandersetzung mit dem Konzept der Mikrostrukturen im Kontext der Modellierung einer lexikografischen Datenbasis. In: Lexicographica 22, S. 72-94.
- Müller-Spitzer, Carolin (2007d): Vernetzungsstrukturen lexikografischer Daten und ihre XML-basierte Modellierung. In: Hermes 38, S. 137-171.
- Müller-Spitzer, Carolin (2010): OWID – A dictionary net for corpus-based lexicography of contemporary German. In: Dykstra, Anne/Schoonheim, Tanneke (Hg.): Proceedings of the XIV EURALEX International Congress (Leeuwarden, 6–10 July 2010). Ljouwert (Leeuwarden), S. 445–452.

- Müller-Spitzer, Carolin/Schneider, Roman (2009): Ein XML-basiertes Datenbanksystem für digitale Wörterbücher – Ein Werkstattbericht aus dem Institut für Deutsche Sprache. In: *it-Information Technology* 51, 4, S. 197-206.
- Sperberg-McQueen, Carl M./Burnard, Lou (1995): The design of the TEI encoding scheme. In: Ide, Nancy/Véronis, Jean (Hg.): *Text Encoding Initiative. Background and context*. Dordrecht, S. 17-39.
- Storjohann, Petra/Möhrs, Christine (2007): Das *ellexiko*-Wörterbuch im Wandel – Typische Verwendungen neu beleuchtet. In: *Sprachreport* 3, 2007, S. 12-16.
- Wiegand, Herbert Ernst (1998): *Wörterbuchforschung. Untersuchungen zur Wörterbuchbenutzung, zur Theorie, Geschichte, Kritik und Automatisierung der Lexikographie*. 1. Teilbd. Berlin/New York.

9.2 Internetressourcen

- OWID – Online-Wortschatz-Informationssystem Deutsch (2008ff.). Institut für Deutsche Sprache (Hg.). Mannheim. Internet: www.owid.de (Stand: 30.04.2010).
- XPath = Clark, James/DeRose, Steve (Hg.) (1999): *XPath — XML Path Language (XPath) Version 1.0*. World Wide Web Consortium, W3C Recommendation 16. November 1999. Internet: www.w3.org/TR/xpath. Deutsche, kommentierte Übersetzung. Internet: www.edition-w3c.de/TR/1999/REC-xpath-19991116 (beide Stand: 30.04.2010).
- XSL = Berglund, Anders (Hg.) (2006): *Extensible Stylesheet Language Version 1.1*. World Wide Web Consortium, W3C Recommendation 05. Dezember 2006. Internet: www.w3.org/TR/xsl/ (Stand: 30.04.2010).
- XSLT = Clark, James (Hg.) (1999): *XSL Transformations (XSLT) Version 1.0*. World Wide Web Consortium, W3C Recommendation 16. November 1999. Internet: www.w3.org/TR/xslt. Deutsche Übersetzung. Internet: www.edition-w3c.de/TR/1999/REC-xslt-19991116 (beide Stand: 30.04.2010).